Response to review of **“Evaluation of CMIP6 Models Performance in Simulating Historical Biogeochemistry across Southern South China Sea”**

Manuscript egusphere-2024-72

Response to Anonymous Referee #2:

We sincerely appreciate the time and effort invested by both the reviewer and the editor in evaluating our paper titled **"Evaluation of CMIP6 Models Performance in Simulating Historical Biogeochemistry across Southern South China Sea"** submitted for publication in Biogeosciences. We are grateful for the positive feedback and the insightful comments provided, which is detailed in this report and also in the upcoming revised manuscript. The majority of the suggestions put forth by the reviewers have been incorporated, and in the limited cases where we have not, we have provided a detailed description of the justification for each decision. For ease of reference, we have provided a detailed point-by-point response to the reviewer’s comments, with line numbers in each response refers to the revised manuscript. Additionally, we have included the recently added discussions, revised figures and newly added references in this “response to the review” report for the convenience of the reviewer/editor to refer.

REVIEWER-2 GENERAL COMMENTS

Comment 1A:

The aim of the paper is to rank 13 CMIP6 ESM simulations based on their ability to reproduce selected observed biogeochemical variables. However, the dataset that the author chose is not strictly observations. Based on the link they provided in line 123, the CMEMS ocean biogeochemistry product is based on the PISCES model output (although it is forced with reanalysis product). I also noticed that among the 13 CMIP6 ESMs, the authors have not chosen IPSL-CM6A ESM, which includes PISCES as its ocean biogeochemical model. I understand that in-situ observations may be rare in this region, but to truly assess the CMIP6 ensemble and individual models, I suggest the authors could compare the CMIP6 models with satellite-derived chlorophyll-a and primary production, as well as the World Ocean Atlas product for nitrate and oxygen.

Response:

- We sincerely appreciate your attention to the matter regarding the usage of reference data (CMEMS) in our study. We understand the importance of ensuring the robustness and reliability of the data utilized in our research, but CMEMS is the only available timeseries hindcast data for the biogeochemistry in southern South China Sea region. We have acknowledged this in our manuscript **(L136)** and emphasized that the quality of CMEMS biogeochemistry product has been validated by Mercator-Ocean. Their validation process, outlined in the Quality Information Document (QuID; Perruche et al., 2019), includes comparisons with recognized datasets such as Ocean Color, World Ocean Atlas, and Globcolour products, ensuring the credibility of the data.
- Furthermore, to boost confidence in this dataset within the southern South China Sea region, we have discussed some literatures in our revised manuscript **(L141 – L150)**, in which, authors have validated this data product with the in-situ measurement in this study region, i.e., Wahyudi et al, (2023) validated the POC, Chlorophyll, Dissolved Oxygen, Nitrate, Phosphate and Silicate obtained from CMEMS biogeochemistry product by comparing it with in-situ data collected during the Widya Nusantara Expedition 2015 (Triana et al., 2021) in the upwelling area of southwestern Sumatra waters. They found that the mean absolute percentage error values were lower than 15%, indicating the reliability of the CMEMS biogeochemistry model data in our study area. Additionally, Chen et al, (2023) also used the daily chlorophyll concentration data from the same CMEMS biogeochemical product in south china sea region. By utilizing this CMEMS biogeochemistry model dataset, Wahyudi et al. (2023) and Chen et al. (2023) highlights the proficiency of the CMEMS biogeochemistry model data in reproducing both the climatic patterns and fluctuations observed within its biogeochemical variables in southern South China Sea. This gave us confidence in utilizing the CMEMS biogeochemical dataset as the reference model to assess other models in this region (southern South China Sea).
- While we appreciate your suggestion of alternative datasets, our decision to utilize CMEMS as the sole reference dataset was made to maintain consistency in the evaluation process. By adhering to a homogeneous dataset, we aim to ensure the integrity and reliability of our evaluation results, thereby instilling greater confidence in our findings.
- Additionally, IPSL-based models showed a standard deviation >50 mg/m^3 for the chlorophyll variable compared to reference data, resulting in their exclusion from the analysis.

Comment 1B:

Since the paper also looks at the seasonal trend of biogeochemical properties, it could benefit from exploring whether different CMIP6 models can capture phytoplankton phenology (e.g., Racault et al., 2015; Gittings et al., 2018), which is an important indicator.

Response:

We appreciate your insightful comment and apologize for any confusion regarding our approach. Our examination primarily focused on the seasonal spatial climatology (now presented as seasonal spatial bias in **Figs. 2-25**), not the seasonal trend map.

While we acknowledge the significance of exploring phytoplankton phenology in CMIP6 models, as suggested by the references you provided, conducting such studies requires extensive time and resources and also, we afraid that incorporating phenology in this study could potentially diverge the main scope of our current investigation. However, we recognize the importance of this aspect and have duly noted it as a potential avenue for future research in our study **(L454 – L458)**. Thank you for bringing this to our attention.

****

Comment 2:

Indeed, most of the biological activity occurs near the surface layers of the ocean, but it's important to consider the biogeochemical dynamics near the seabed, particularly in shelf seas, as they can have complex structures through interactions of ocean physics with biological processes, such as export and remineralization. I would appreciate the inclusion of depth profiles and benthic concentrations of oxygen and nitrate – this would provide a more thorough assessment of the biogeochemical properties. Furthermore, most of the biogeochemical models used in CMIP6 are not specifically built for shelf seas. It would be interesting to see whether these models can represent nutrient and oxygen distribution at shallower depths.

Response:

- Thank you for bringing up the importance of analysing the bias at depth to understand the oxycline and nutricline dynamics of the models. In response, considering the complex bathymetry of southern SCS region, we have addressed this concern by presenting the spatial distribution of seasonal variations in nitrate and oxygen at two distinct depths (70m and 1000m) for each model, rather than providing profiles **(L270 – L320)**. The depth of 70 meters has been selected to depict the dynamics of the nutricline/oxycline in the shelf break region of Sunda Shelf region, while a depth of 1000 meters has been chosen to represent the deep layer. Accordingly, we have discussed the biases

- For nitrate **(L278 – L291)** as follows:

  "Furthermore, delving into model biases at deeper levels, especially concerning nutrient dynamics, will provide more insights into the model's accuracy in simulating the nutricline. Consequently, we analysed the nitrate concentrations at depths of 70m ***(Figs. 11 – 13)*** *and 1000m* ***(Figs. 14 - 16)****. In contrast to surface nitrate, most models exhibited a negative bias at deeper layers (70m and 1000m), with an average range of -2 to -8 mmol/m^3 across the study area. Among these models, MPI-based models showed the least negative bias at 70m depth; however, as depth increased to 1000m, their bias shifted towards the positive* ***(Fig. 13 and 16****, respectively***)***. MIROC-based and MPI-based models exhibited the least bias in nitrate concentrations at both surface and deep layers compared to reference data. This may be attributed to the near balance achieved between nitrogen cycle sources (such as nitrogen fixation, atmospheric nitrogen deposition, and riverine nitrogen input) and sinks (including denitrification, nitrous oxide emission, and sedimentary loss) over the long spin-up period (Mauritsen et al., 2019; Hajima et al., 2020). In contrast, CanESM5-based models demonstrated minimal nitrate bias at the surface but showed varying positive and negative biases in deep layers. These discrepancies arise from the simplified parameterization of denitrification in their BGC models. In these models, denitrification in the deep layers is set to balance the rate of nitrogen fixation and is vertically distributed in proportion to the detrital remineralization rate. However, in reality, nitrogen fixation and denitrification are not constrained to balance within the water column at any single location; rather, denitrification primarily occurs in anoxic areas (Swart et al., 2019). Notably, no seasonal bias in all selected models were observed at the deep layer (1000m;* ***Fig. 16****)."*

- For oxygen **(L293 – L313)** as follows:

  "During the observation of oxycline dynamics in the selected models, it was noted that the oxygen exhibited a positive bias at a depth of 70m, transitioning to a negative bias with increasing depth (1000m)

(Fig. 20 and 25). Moreover, UKESM1-0-LL consistently exhibited a substantial positive bias from the surface to the depth of 70m (~50 mmol/m³) and shifts to negative bias of -40 mmol/m³ at 1000m relative to its surface bias. Similarly, CanESM5 and MIROC-based models displayed markedly high negative biases at a depth of 1000m, but with comparatively lesser negative biases at 70m. Multiple factors could contribute to biases in the simulation of nutricline/oxycline dynamics by models. Inaccuracies in simulating nutricline dynamics may arise from errors in parameterizing physical, chemical and biological processes relevant to these dynamics. In winter, most models overestimate oxygen levels at the surface and at a depth of 70 meters. This positive bias in oxygen concentration may result from excessively intense winter mixing of cold, oxygen-rich waters from the northern boundary of the southern SCS into the Sunda Shelf region (Thompson et al., 2016), which transports an excessive amount of surface oxygen to deeper layers. Additionally, nutrient trapping issues may also contribute to the remaining model bias (Six & Maier-Reimer, 1996). Moreover, the exclusion of relevant processes or feedback mechanisms influencing nutricline dynamics within the model, such as nutrient upwelling, microbial remineralization and ocean stratification, may lead to biased simulation outcomes. Additionally, structural uncertainties embedded in the model formulation, including simplifications or assumptions regarding complex processes, may also play a role in generating biases in simulation results. For example, advancements in model parameterization and representation of biogeochemical fluxes have led to consistent improvements in the mean states of nutrient dynamics in CMIP6 models, such as GFDL-ESM4, MIROC-based, MPI-ESM1-based, and NorESM2-based models. Specifically, improvements in GFDL-ESM4 performance are attributed to a series of updates and changes in model physics (such as mixing and climate dynamics) and biogeochemical parameterizations such as the implementation of a revised remineralization scheme for organic matter that depends on oxygen and temperature (Laufkötter et al., 2017)."

****

Comment 3:

Although the authors put a great effort in evaluating CMIP6 model outputs, the model structures could also be evaluated; how biogeochemical tracers are represented, and whether these representations affect the performance of the model in the southern SCS. Perhaps the authors can add another table which biogeochemical tracers these models represent (e.g., in MEDUSA-2 (UKESM), it does not represent diazotrophic phytoplankton, explicitly calculates phytoplankton chlorophyll, and uses N as model currency, while in OECO-2 (MIROC), it has diazotrophic phytoplankton with C as model currency and includes Phosphate as nutrients), and perhaps also how they are formulated, especially when it involves trophic transfer (e.g. nutrient uptake, zooplankton grazing, and phytoplankton growth, and plankton mortality). These additions can add some discussion on how model representation (and structure) may affect model performance in the shelf seas, instead of repeatedly saying that underestimation/overestimation is due to zooplankton grazing/phytoplankton productivity/nutrient uptake.

Response:

We sincerely appreciate your insightful comments regarding the model structures. Following your suggestion, we have incorporated an overview of how tracers and model structure affect performance into our discussion, specifically when examining inter-parameter relationships such as chlorophyll-biomass and biomass-nitrate in in **L432 – L454** as follows:

"This slight negative correlation could stem from various factors that it may reflect discrepancies in those model dynamics, such as the representation of nutrient uptake or phytoplankton growth rates. Biological processes within the models might not accurately capture the complexities of phytoplankton-nutrient interactions. For example, variations in biogeochemical tracers within model frameworks could influence model efficacy. Specifically, except UKESM1-0-LL and MIROC-based models, all other selected models utilize carbon as their primary model currency for representing phytoplankton biomass, incorporating explicit calculations for phytoplankton biomass and they also utilize nitrate and phosphate to constrain bulk phytoplankton growth rates alongside temperature and light. Consequently, their representation of phytoplankton biomass exhibited a weaker correlation with nitrate. Despite the use of carbon tracer, MPI-ESM1-2-LR incorporates a newly resolved nitrogen-fixing formulation within its biogeochemistry model. This updated formulation introduces an additional

prognostic phytoplankton class, replacing the diagnostic formulation of nitrogen-fixation utilized in MPI-ESM-LR (Paulsen et al., 2017; Mauritsen et al., 2019). As a result, this adjustment enables the model to capture the nitrogen response to phytoplankton biomass positively. UKESM1-0-LL employed nitrogen as its primary currency, resulting in a more pronounced quantitative representation of phytoplankton biomass in response to increased nitrate levels compared to the other models ***(Fig. 27)****. While MIROC-ES2L primarily utilizes nitrogen as its tracer, it also integrates the phosphorus cycle within the model framework to accurately depict the strong phosphorus limitation on the growth of diazotrophic phytoplankton (Hajima et al., 2020). Consequently, this incorporation of the phosphorus cycle may account for phosphorus limitation, resulting in the observed negative correlation between nitrate and phytoplankton biomass within our study area. In the case of GFDL-ESM4, the negative correlation between nitrate and phytoplankton could potentially originate from their model parametrization. In their framework, phytoplankton were categorized based on size and functional type, with small phytoplankton being nitrogen-rich and large phytoplankton phosphate-rich, thereby attributing characteristic N:P ratios (Stock et al., 2020). Thus, differences in parameterizations, data initialization and model resolution could contribute to divergent simulated responses (Séférian et al., 2020)."*

****

Comment 4:

The presentation of the results can also be improved. I think it will be easier to follow the results if the authors describe the observed distribution of nitrate, chlorophyll, phytoplankton biomass, and oxygen, then compare them with the model. For the figures, it would be more interesting to see the difference between the CMEMS data and CMIP6 outputs with better figure resolution (especially figure 6). Additional discussion on regions where bias usually occurs in different models will also be interesting (e.g., the shelf seas between Sumatra, the Malaysian peninsula, and Borneo are always high in phytoplankton biomass for UKESM, CanESM5, ACCESS, MPI-ESM1-2, NorESM2).

Response:

- We sincerely appreciate your insightful feedback and acknowledge your concern regarding the clarity of the results. Following your suggestion, we have modified the seasonal climatology figure in our revised manuscript to better illustrate the seasonal bias against the reference data **(Figs. 2 – 25**: can be found in this report below**)**.
- Additionally, we have incorporated your recommendation to discuss the spatial diversity in model bias. Our revised manuscript now elaborates this in **L235 – L250** as follows:

  "Most models showed spatial uniformity in their underestimation or overestimation of chlorophyll and phytoplankton, with a few models exhibiting spatial diversities in their estimates. For example, the CanESM-CanOE model consistently overestimates chlorophyll concentration and phytoplankton biomass in both seasons, with mean biases of ~0.49 mg/m³ in DJF and ~0.31 mg/m³ in JJA for chlorophyll, and ~2.2 mmol/m³ in DJF and ~1.5 mmol/m³ in JJA for phytoplankton. This overestimation is particularly pronounced in the region between Sumatra, Peninsular Malaysia, and Borneo, where chlorophyll exceeds 1 mg/m³ and phytoplankton exceeds 5 mmol/m³. Similarly, the UKESM1-0-LL model overestimates chlorophyll and phytoplankton in both seasons, especially in the Gulf of Thailand. These models may have insufficient spatial resolution to capture the fine-scale physical and biological processes in these regions. Important features like small-scale currents, eddies and upwelling events, which significantly affect chlorophyll and phytoplankton distributions, may not be adequately resolved, leading to spatial bias. Generally, ESMs in CMIP6 are developed for open ocean conditions rather than shelf seas. Most CMIP6 ESMs have a coarse resolution (≥ 100 km horizontal) for the ocean component, although some have resolutions of ≤ 25 km, which is considered eddy-permitting. This suggests these ESMs can represent barotropic processes at smaller scales but not baroclinic ones (Chelton et al., 1998). The ability of coarse-resolution CMIP6 ESMs to represent shallow continental shelf waters dynamics with high skill, such as the southern SCS' Sunda shelf region, is limited. Variability in this region is influenced by inflows like the Indonesian Throughflow and SCS Throughflow, which are not resolved by coarse-resolution models (Wang et al., 2024)."

****

REVIEWER-2 SPECIFIC COMMENTS

Comment 1:

L12 – perhaps the authors can add a % or number on the degrees of overestimation and underestimation.

Response:
Thank you for highlighting this important point. We apologize for the oversight and any confusion it may have caused. The overestimations or underestimations pertain to quantitative measures that vary depending on the analysed variables and seasons. We have included this clarification in the revised manuscript, following your recommendation **(L13)**.

****

Comment 2:

L22-23 - Based on CMIP6 models, NPP trend is uncertain, apart maybe at the Southern Ocean (Tagliabue et al., 2021)

Response:
Thank you for your valuable feedback. We acknowledge the inaccuracy in our original statement regarding marine NPP uncertainty in 2100. Accordingly, the corrections have been made in revised manuscript **(L23 – L29)** as follows:

"For example, Kwiatkowski et al. (2020) discovered that the multi-model global mean projections from the Coupled Model Intercomparison Project Phase 6 (CMIP6), under high-emission to low-emission scenarios, indicate a consistent decrease in net primary production. Notably, there is a significant increase in inter-model uncertainty compared to CMIP5. This increased uncertainty is linked to changes in the temporal patterns of phytoplankton resource availability and grazing pressure within CMIP6 (Kwiatkowski et al., 2020). This carries significant implications for evaluating ecosystem impacts on a regional scale. (Tagliabue et al., 2021)."

****

Comment 3:

L33-34 - This is not always the case - OBGC models can give the seemingly good representation of historical climate pattern but for the wrong reason. Furthermore, OBGC model results is dependent on its physical forcings (see Sinha et al., 2010)

Response:
Thank you for bringing this matter to our attention. In response to your suggestion, we have gone through some literatures and made changes accordingly in **L35 – L42** stating that *"While a model's successful reproduction of historical climate patterns suggests it has captured relevant physical processes and interactions within the Earth's system but this is not always the case. Ocean BGC models can sometimes appear to accurately represent historical climate patterns for incorrect reasons, as their results are highly dependent on the physical forcing applied (Friedrichs et al., 2006; Sinha et al., 2010). For example, minor changes in ocean model circulation can lead to significant variations in biogeochemical conditions. Similarly, Glessmer et al. (2008) discovered that even slight alterations in mixing greatly affects the simulation of primary production and export in the global general circulation models. Therefore, caution is needed when using these models for future climate projections."*

****

Comment 4:

L60 – typo: Tjiputra et **al**, (2020)

Response:
Thank you for pointing out this oversight. We have corrected accordingly in **L68**

****

Comment 5:

L61-72 – I'm not so sure if these are appropriate examples. Maybe add studies like Kwiatkowski et al., 2020, Hinrichs et al., 2023

Response:

Thank you for your valuable suggestion and for providing literature on this topic. We have thoroughly reviewed the references you recommended, and we have incorporated relevant findings from Kwiatkowski et al., 2020 into our manuscript **L88**.

****

Comment 6:

L83-L85 – Why only phytoplankton, chlorophyll, nitrogen, and oxygen? Why not net primary production and or carbon?

Response:

Thank you for your concern about the selected variables. We chose to focus on phytoplankton, chlorophyll, nitrogen, and oxygen for following reasons:

1. These variables are fundamental tracers for biological and nutrient dynamics in ocean systems.
2. Phytoplankton and chlorophyll variables serve as effective proxies for primary production, thus encompassing the essential aspect of net primary production.
3. Given that the southern SCS region is recognized as a typical oligotrophic area where primary productivity is primarily constrained by nutrient availability, we specifically included nitrate and oxygen.
4. We considered variables that were consistently available across all selected models' historical and projection scenarios to ensure comparability and consistency in our analysis. Even phosphate variable is unavailable in some of the selected model's scenarios. Thus, phosphate is also excluded.

****

Comment 7:

L103-L105 - This sounds like phytoplankton is controlling the physical biogeochemical process?

Response:

We sincerely apologize for the confusion caused by our statement. Upon a careful review, we rephrased the statement as ***"****Within the southern SCS, extensive observations have demonstrated that phytoplankton growth, serving as the primary source of organic matter, significantly influences oceanic carbon cycles. This growth is influenced by monsoon-driven physical and biogeochemical processes, with phytoplankton demonstrating a notable sensitivity to these environmental dynamics.* ***"*** in **L124 – L127**.

****

Comment 8:

L122-123 – is this the hindcast global ocean biogeochemistry? Do you also use the GlobColour for chlorophyll? Please be more specific.

Response:

- We sincerely apologize for the confusion made. The CMEMS product used in this study is the hindcast global ocean biogeochemistry dataset, which can be found in CMEMS biogeochemistry hindcast dataset (ID: GLOBAL_MULTIYEAR_BGC_001_029) **(L156).**
- We did not use GlobColour data in our study. Instead, Mercator Ocean used GlobColour data to validate the chlorophyll data within the CMEMS hindcast biogeochemistry dataset, as mentioned in **L138 – L141**.

****

Comment 9:

L125 – Perhaps, instead of having 2/3 ESMs with the same OBGC model, maybe choose one of them instead, so you can also look at other models such as PISCESv2 (Aumont et al., 2016), MARBL (Long et al., 2021), BFM5.2 (Lovato et al., 2022)?

Response:
Thank you for suggesting a method to choose models. We will incorporate this method in our future studies. Additionally, as explained in **L129 – L132** that our current model selection procedure was based on the availability of selected biogeochemical variables across historical or projected scenarios. Based on this, CESM2 model utilizing the MARBL bgc model was excluded from our study due to the absence of the dissolved oxygen (o2) variable in its historical dataset.

****

Comment 10:

L132 – Do you mean visualised using taylor diagram? How do you calculate model/data comparison using a diagram?

Response:
We sincerely apologize for the confusion caused by our statement. Upon a careful review, we replaced the word *"calculated"* to *"visualized"* in **L165**.

****

Comment 11:

L164-165 Can you provide a reference on this statement?

Response:
Thank you for your concern regarding this matter. The statement is nuanced, reflecting both aspects. While some studies, such as those by Behrenfeld et al. (2006) and Kwiatkowski et al. (2017), indicate that seasonal cycles can help constrain projections, our intended message **(L197 – L201)** is that *"although temporal cycles like the yearly seasons are important components of climate variability, they provide only a partial perspective on long-term climate change. Long-term changes involve shifts in average temperatures, changes in precipitation patterns, variations in the frequency and intensity of extreme weather events, and other systemic transformations that go beyond the periodic nature of seasonal cycles."* Thus, our revised statement **(L197 – L201)** aims to convey that "*the yearly cycle of seasons partially captures the long-term changes associated with climate change*".

****

Comment 12:

L172 – but CMEMS data is not really observation, isn't it?

Response:
We sincerely apologize for the confusion caused by our statement. CMEMS is not observation data. We replaced the word *"observed"* to *"reference"* in **L208**.

****

Comment 13:

L177-L179 - perhaps spell out how these models represent their phytoplankton growth and chlorophyll concentration? And compare it to models that have better RMSD?

Response:
Thank you for highlighting this concern. The model's representation of biological tracers was discussed in Inter-variable relations section in **L395** as detailed in **Reviewer-1 major comment 1**.

****

Comment 14:

L184, 219 – what is acceptable range?

Response:
Thank you for highlighting this concern. We have addressed it by representing the acceptable bias range based on models with small mean bias. Accordingly, in the revised manuscript, we have specified the acceptable range as $\leq \pm 0.3$ mg/m³ for chlorophyll in **L220** and $\leq \pm 1$ mmol/m³ for phytoplankton in **L234**.

****

Comment 15:

L186 - why is UKESM not overestimating chlorophyll, but overestimates phytoplankton carbon?

Response:
Thank you for your concern regarding this matter. In **L225 – L229** we explained that *"UKESM1-0-LL model explicitly simulates chlorophyll concentrations, allowing for a more accurate representation of chlorophyll levels (Sellar et al., 2019). However, UKESM1-0-LL uses nitrogen as its primary model currency, which results in a more pronounced quantitative representation of nutrient levels. This might lead to enhanced nutrient uptake by phytoplankton due to differences in model parameterizations and consequently result in the overestimation of phytoplankton biomass."* This could explain why the model does not overestimate chlorophyll but does overestimate phytoplankton.

****

Comment 16:

L232-L242 – maybe move this to the study domain part instead of on the results section?

Response:
Thank you for your suggestion. Accordingly, we shifted this part to study domain section in **L109 – L119**.

****

Comment 17:

L251 – Can you give example of the important processes?

Response:
Thank you for your insightful comment. The important processes that may be overlooked by some ESMs include nutrient cycling, light availability, temperature variations, and phytoplankton phenology. These processes play important roles in shaping the seasonal patterns of biogeochemistry in marine ecosystems. We have clarified this point in the revised manuscript in **L390**.

****

Comment 18:

L286 **-** Why do you think this is? could it be that the ESMs in CMIP6 is developed based on the condition of the open ocean, but not the shelf seas? Or is it because the resolution is too coarse for shelf seas?

Response:
Thank you for your insightful question. Indeed, both factors you mentioned could contribute to the observed performance of the ESMs in simulating biogeochemical variables. The ESMs in CMIP6 are primarily developed based on the conditions of the open ocean, which may not fully capture the complexities of shelf seas. Additionally, the coarse resolution of these models may not adequately resolve the fine-scale processes occurring in shelf sea environments. Together, these factors likely contribute to the moderate to poor performance of the ESMs in simulating biogeochemical variables, as mentioned in **L502 – L506**.

****

Comment 19:

Figure 6 could do with higher resolution.

Response:

Thank you for your suggestion regarding Figure 6. Accordingly, we have resolved the clarity of the Figure, which is now **Fig. 12** (can be found in this report below).

****

Revised Figures:

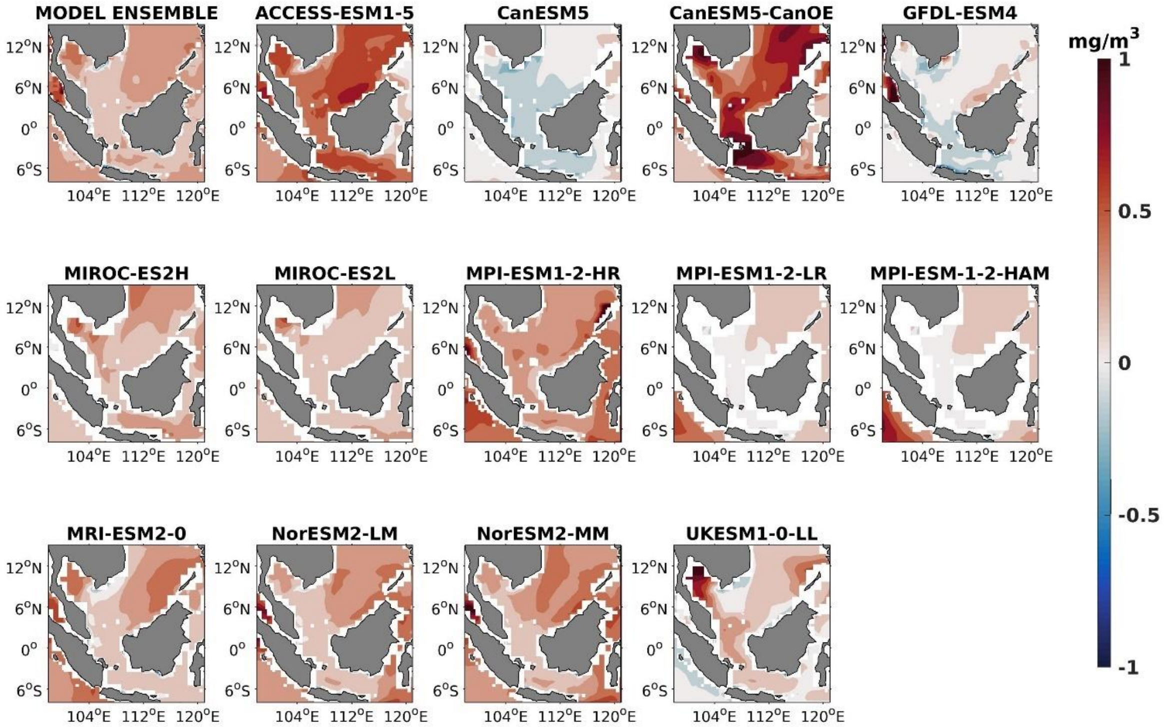

Figure 1. DJF spatial biases of surface chlorophyll for 13 individual models and model ensemble relative to reference.

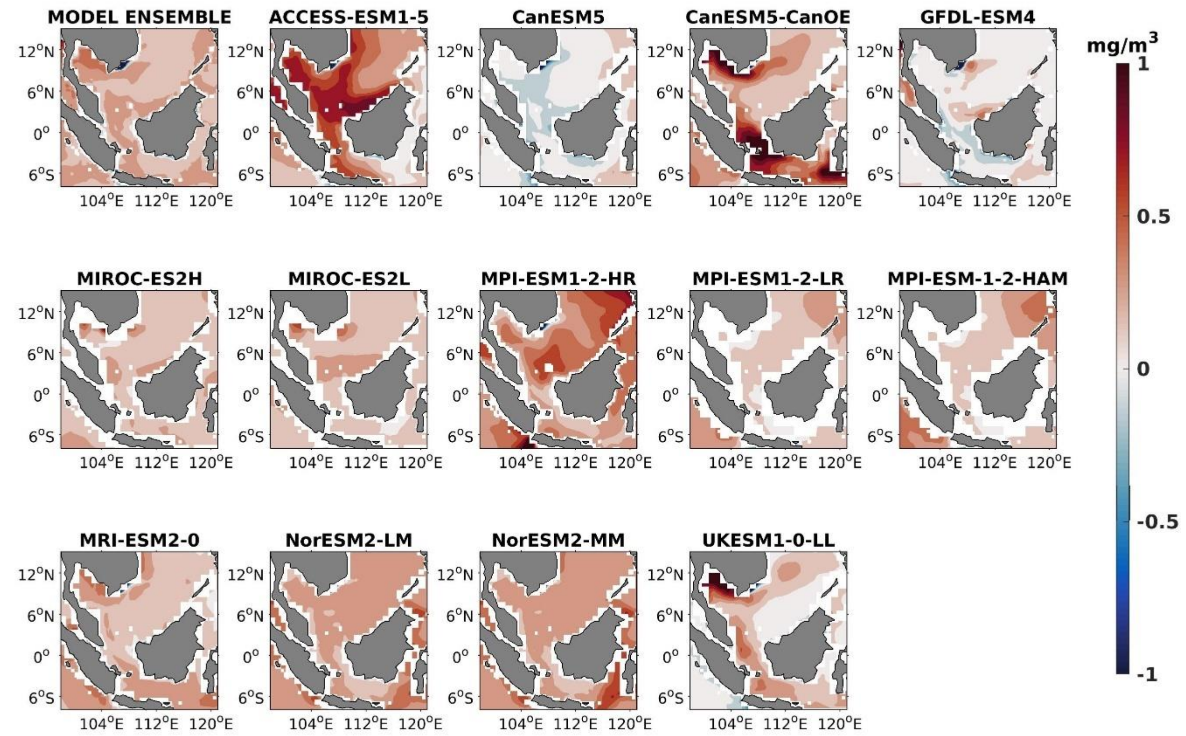

Figure 2. Same as Fig.2 but for JJA.

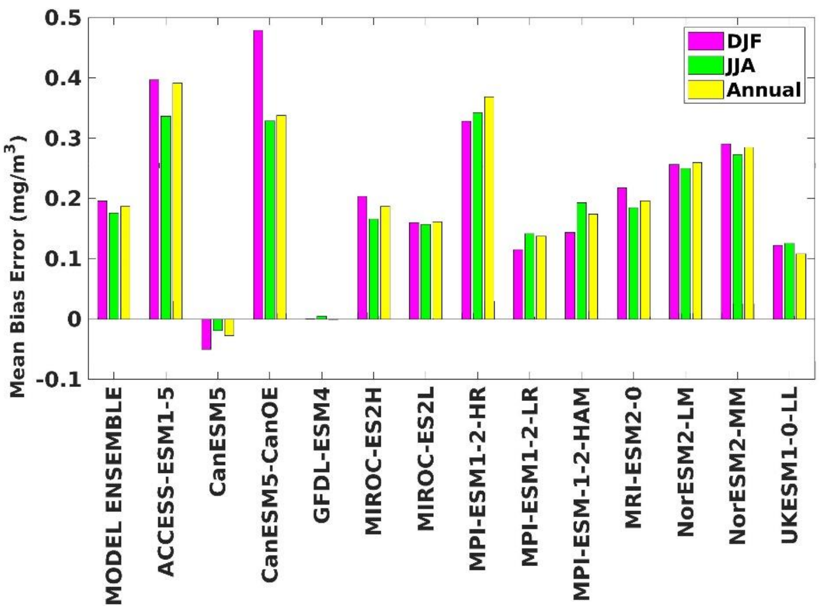

Figure 3. The mean bias of surface chlorophyll for both seasons (DJF, JJA) and annual.

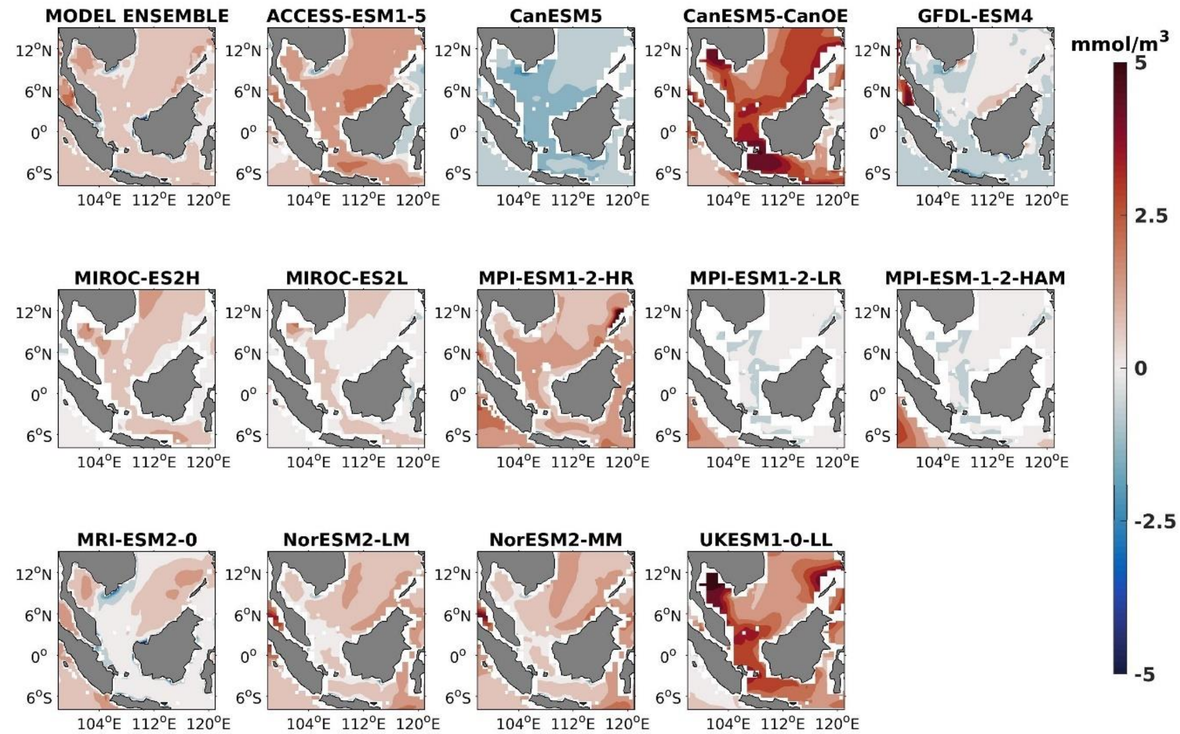

Figure 4. DJF spatial biases of surface phytoplankton for 13 individual models and model ensemble relative to reference.

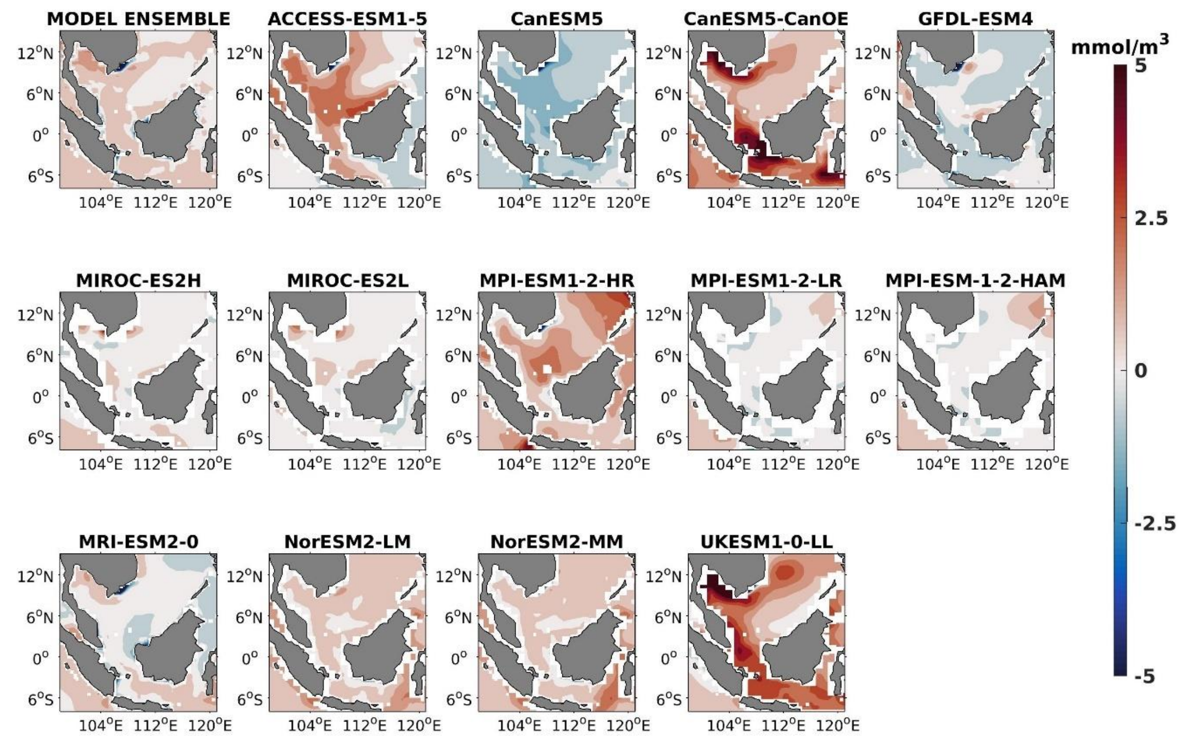

Figure 5. Same as Fig. 5 but for JJA.

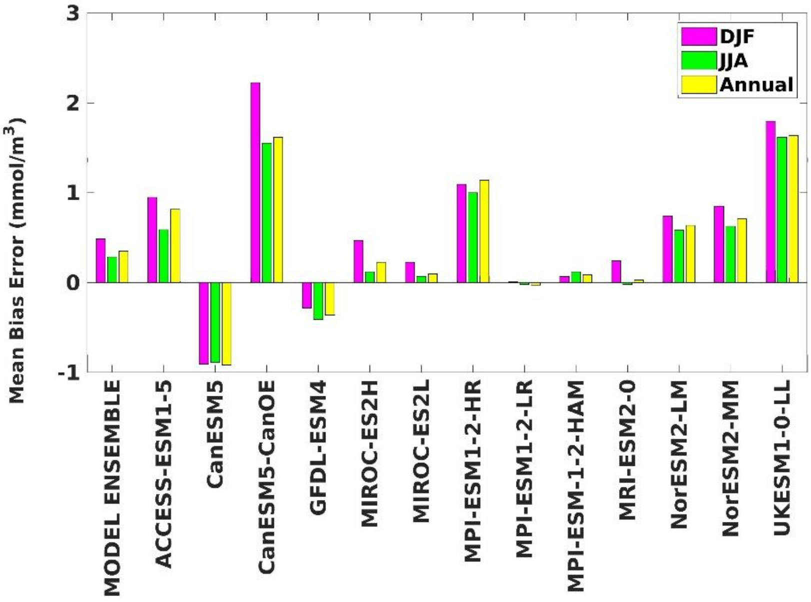

Figure 6. The mean bias of surface phytoplankton for both seasons (DJF, JJA) and annual.

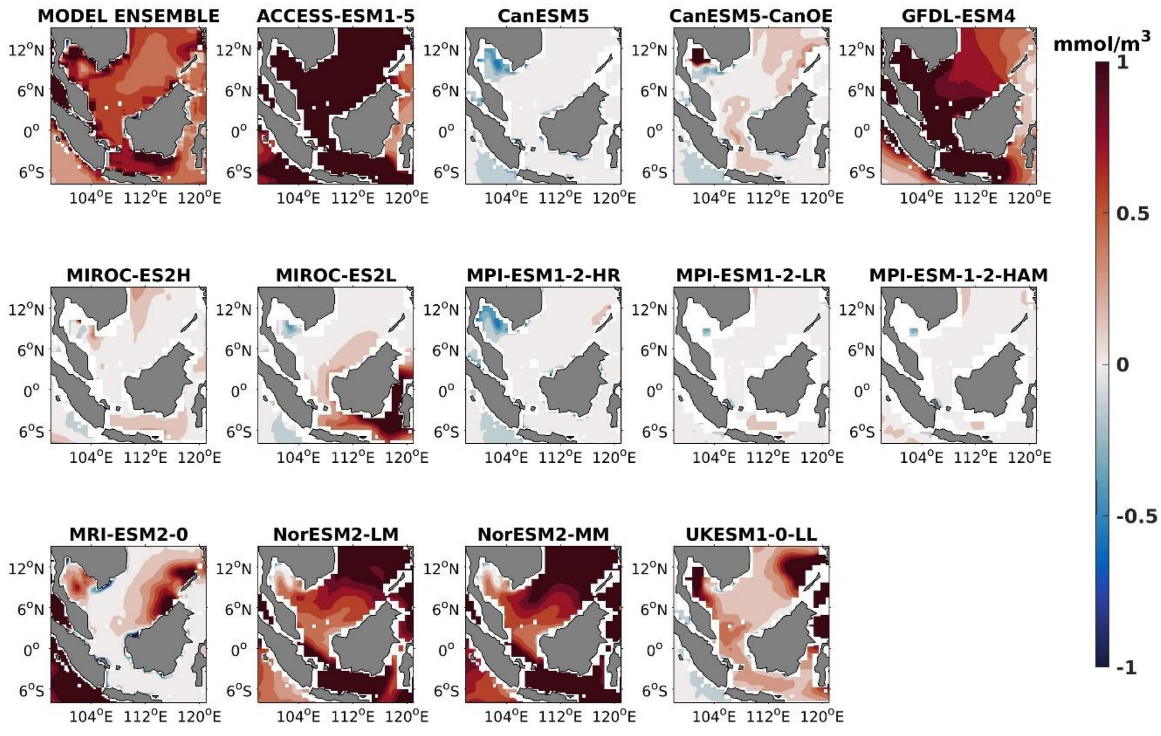

Figure 7. DJF spatial biases of surface nitrate for 13 individual models and model ensemble relative to reference.

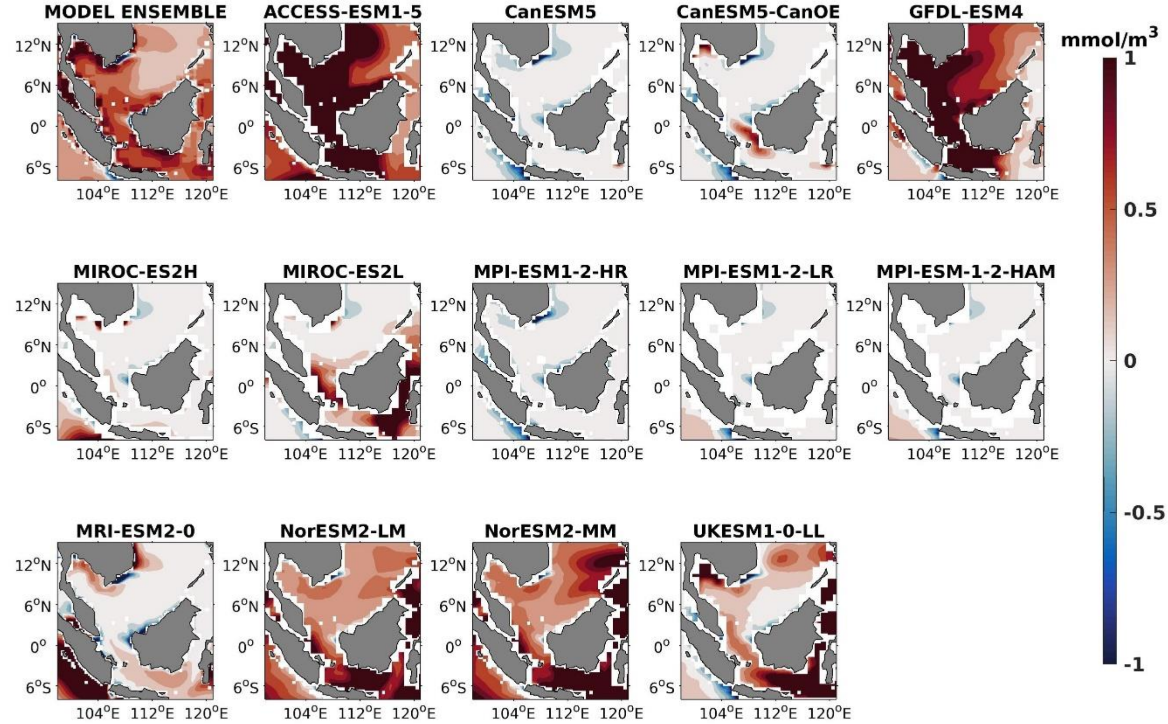

Figure 8. Same as Fig. 8 but for JJA.

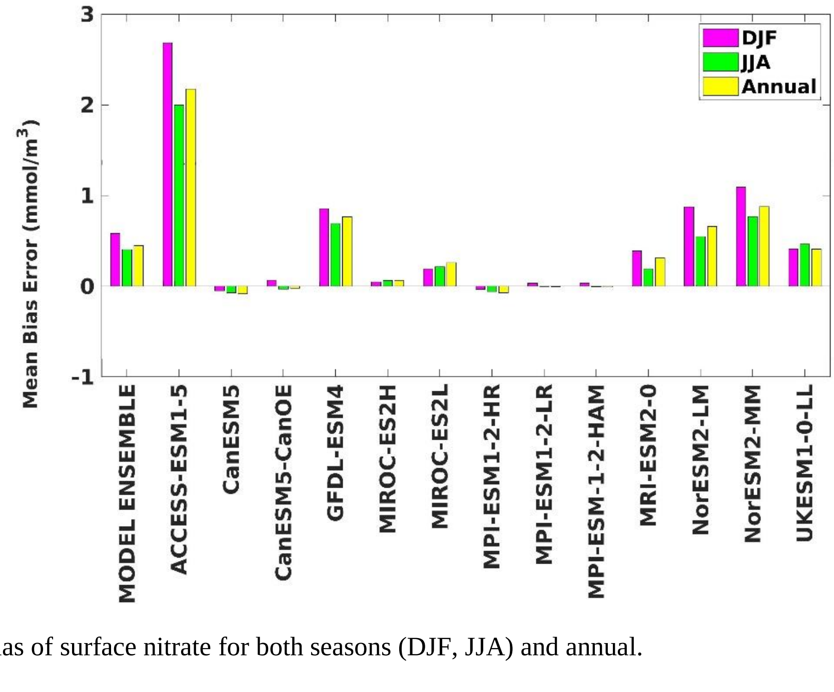

Figure 9. The mean bias of surface nitrate for both seasons (DJF, JJA) and annual.

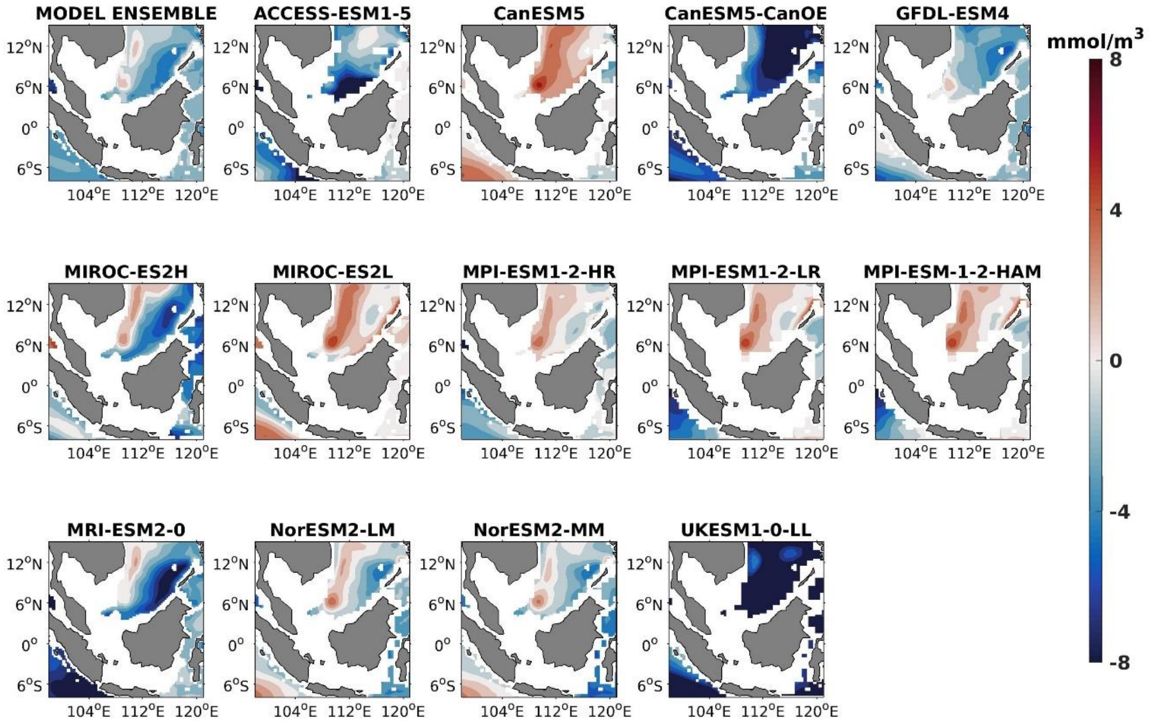

Figure 10. DJF spatial biases of nitrate at 70-meter depth for 13 individual models and model ensemble relative to reference.

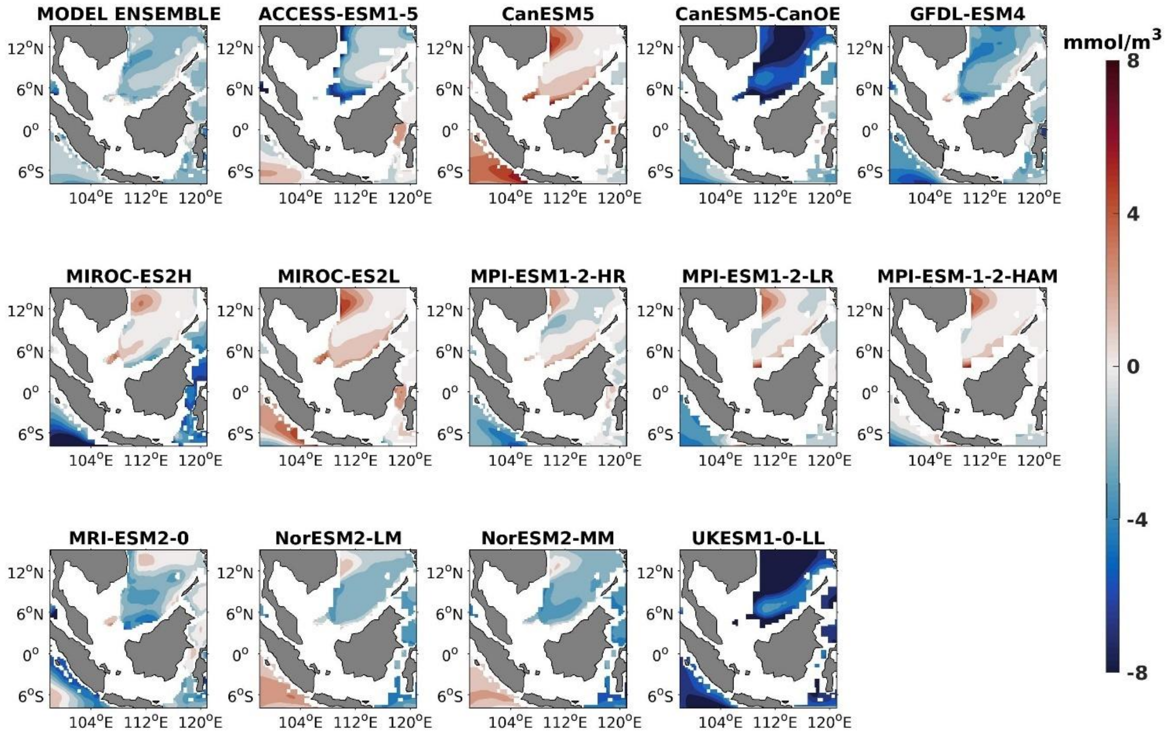

Figure 11. Same as Fig. 11 but for JJA.

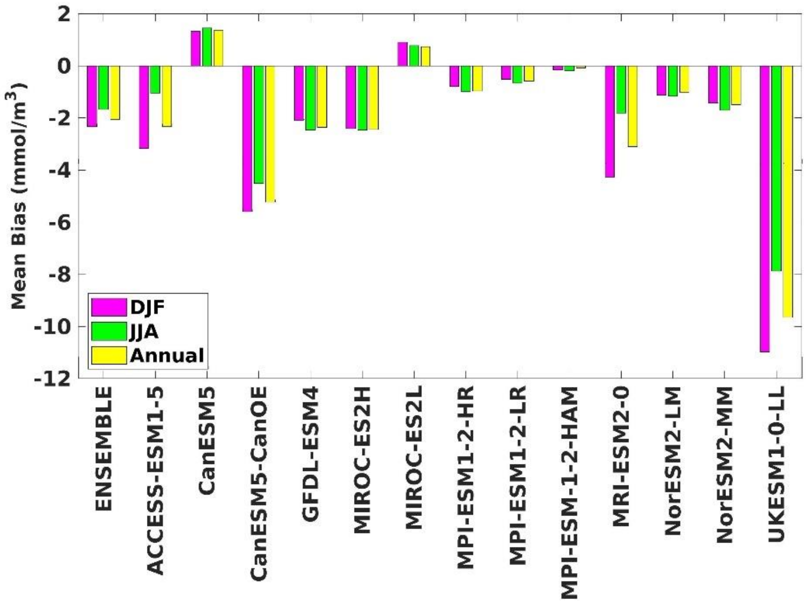

Figure 12. The mean bias of nitrate at 70-meter depth for both seasons (DJF, JJA) and annual.

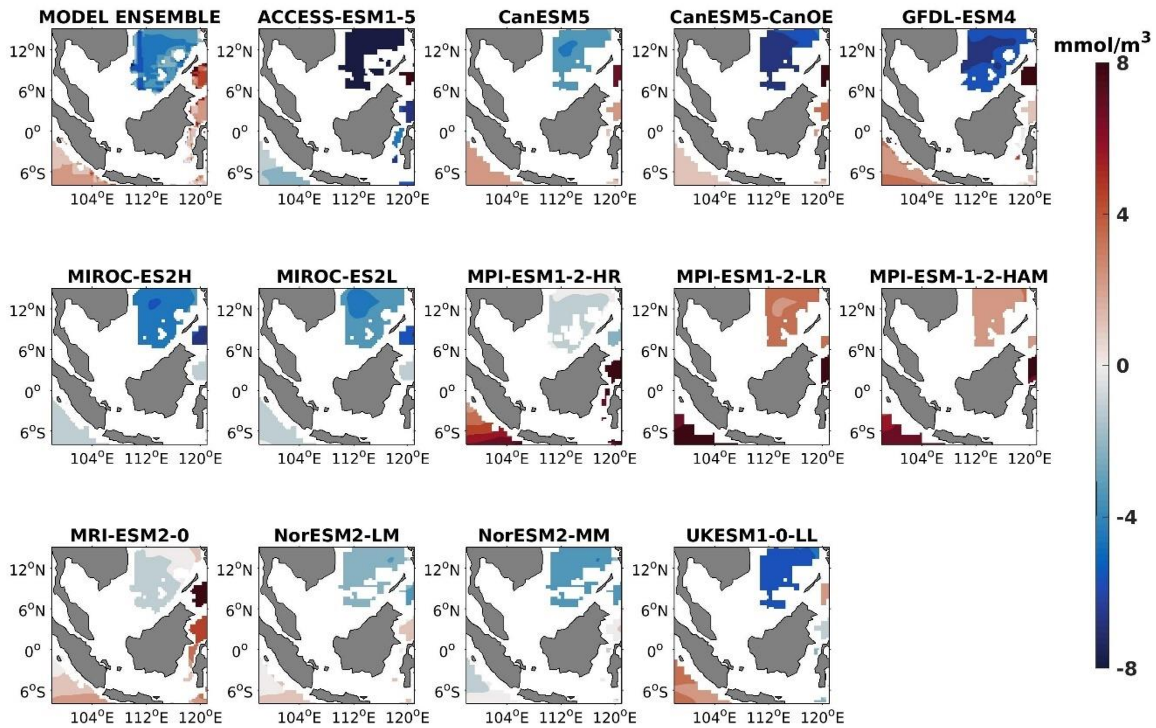

Figure 13. DJF spatial biases of nitrate at 1000-meter depth for 13 individual models and model ensemble relative to reference.

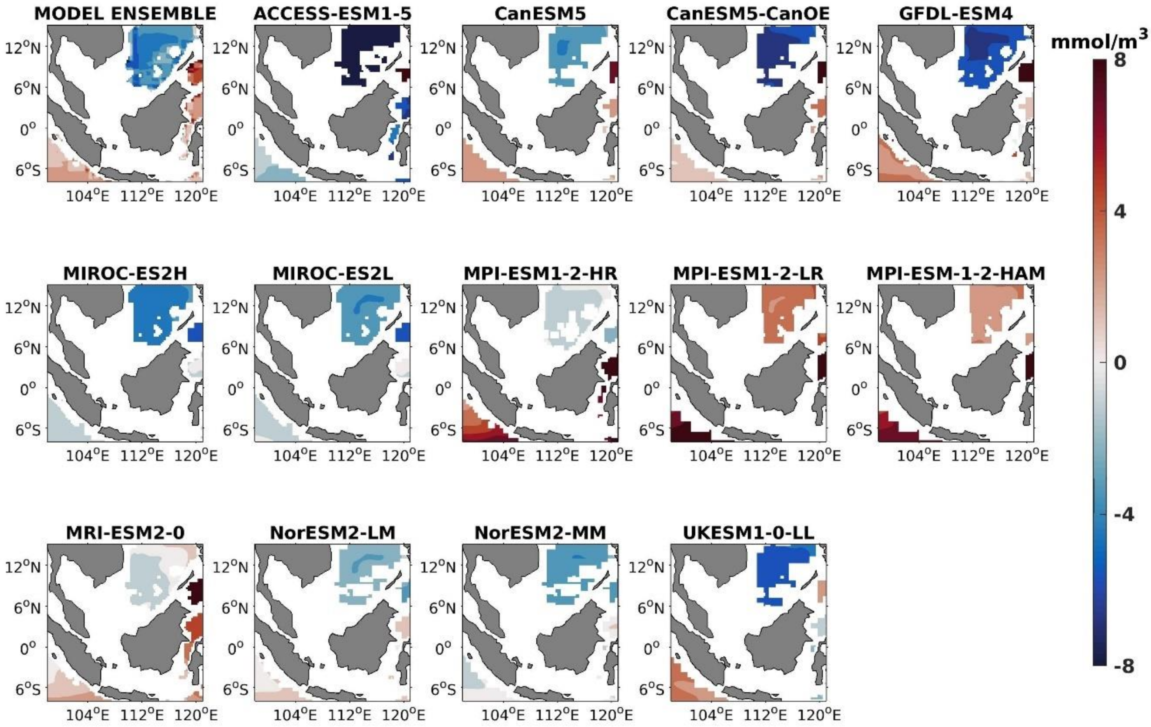

Figure 14. Same as Fig. 14 but for JJA.

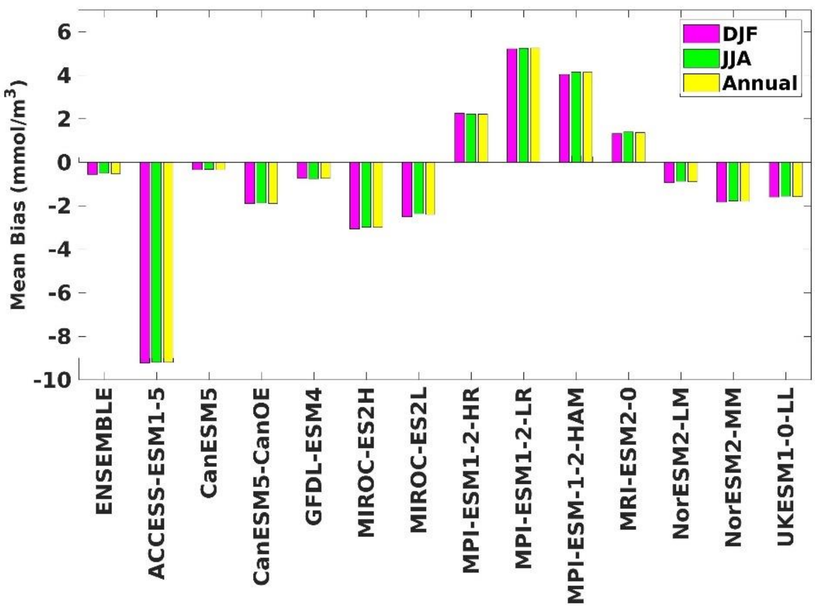

Figure 15. The mean bias of nitrate at 1000-meter depth for both seasons (DJF, JJA) and annual.

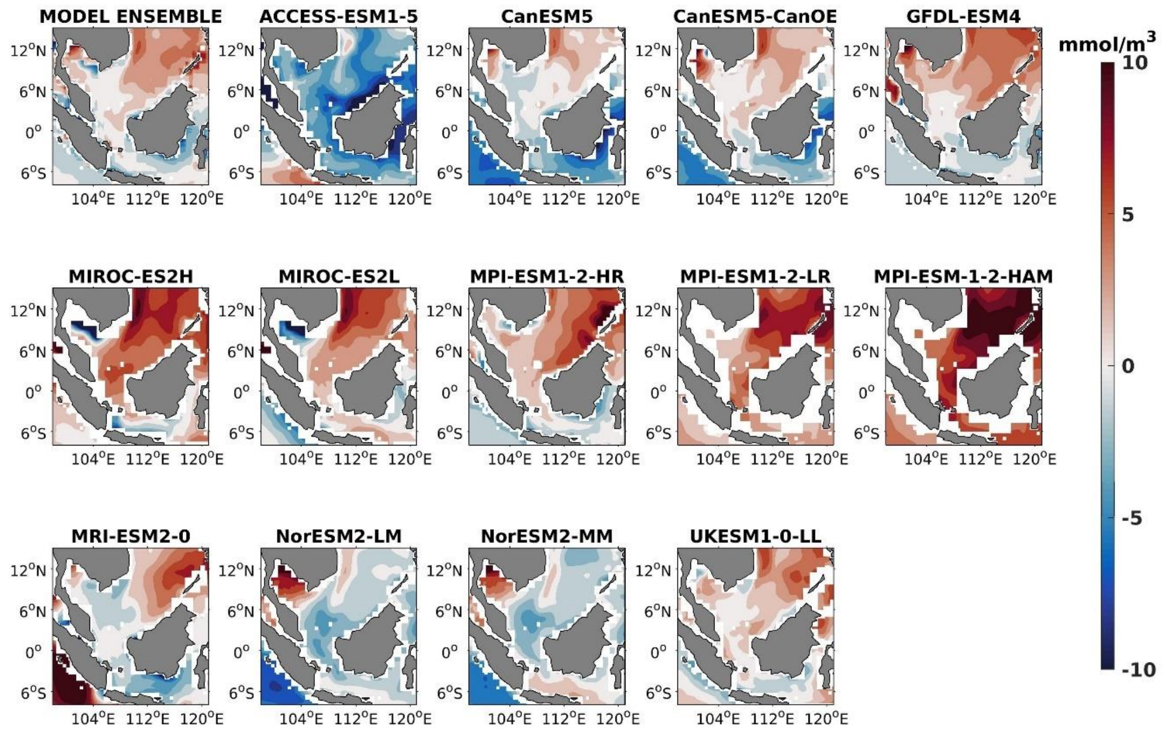

Figure 16. DJF spatial biases of surface oxygen for 13 individual models and model ensemble relative to reference.

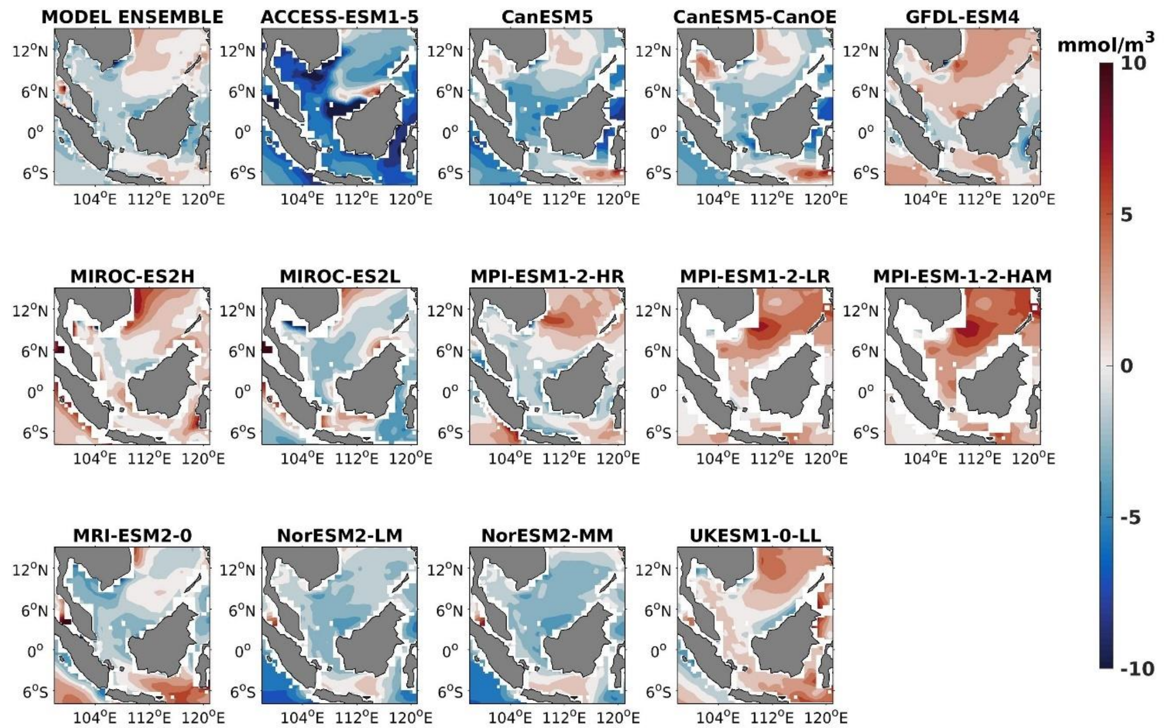

Figure 17. Same as Fig. 17 but for JJA.

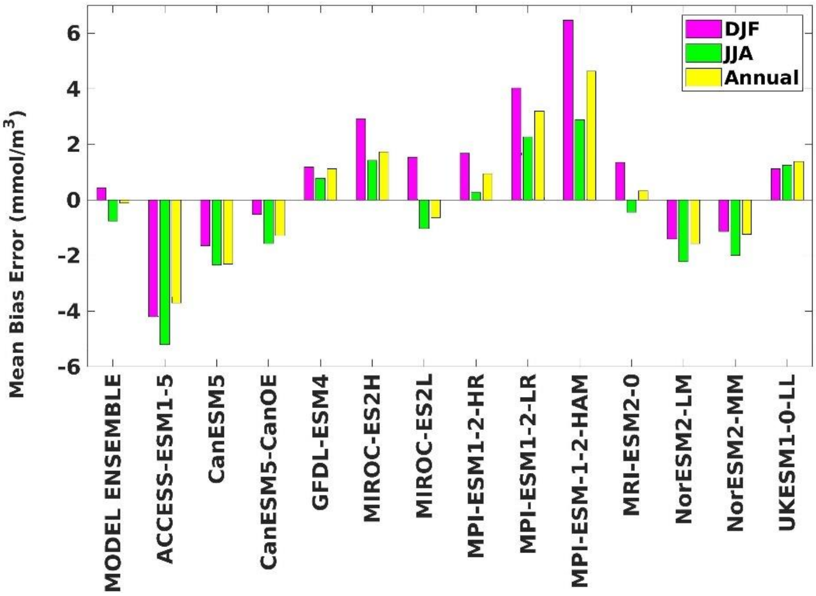

Figure 18. The mean bias of surface oxygen for both seasons (DJF, JJA) and annual.

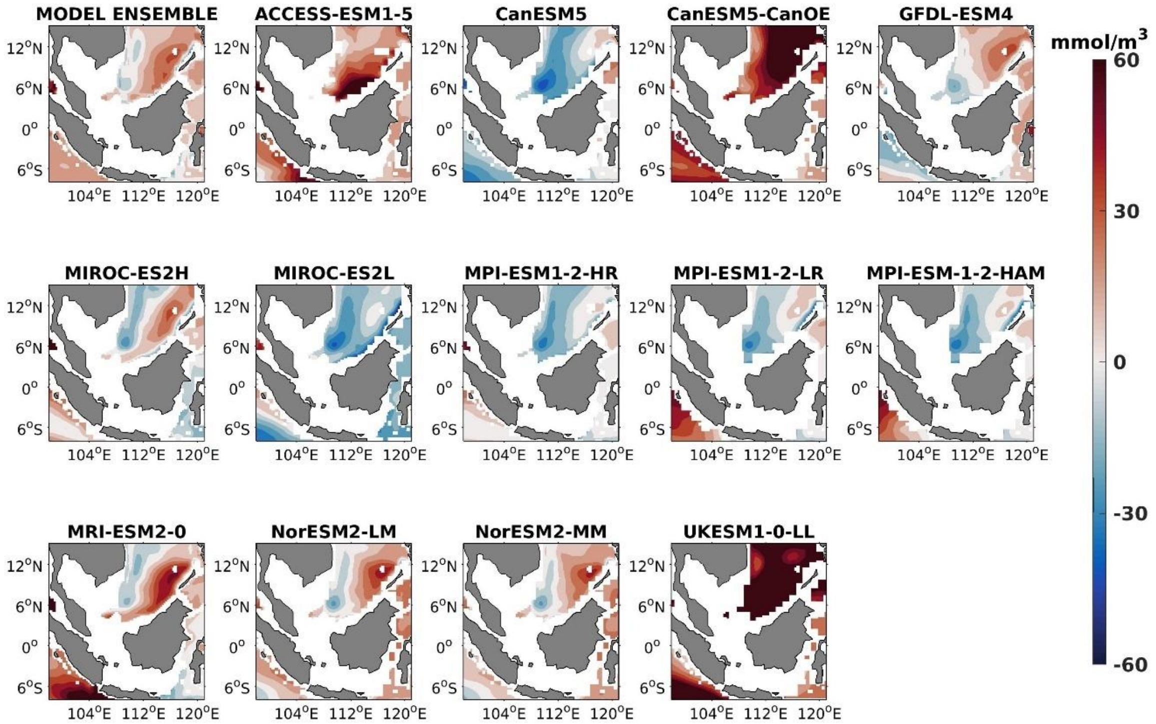

Figure 19. DJF spatial biases of oxygen at 70-meter depth for 13 individual models and model ensemble relative to reference.

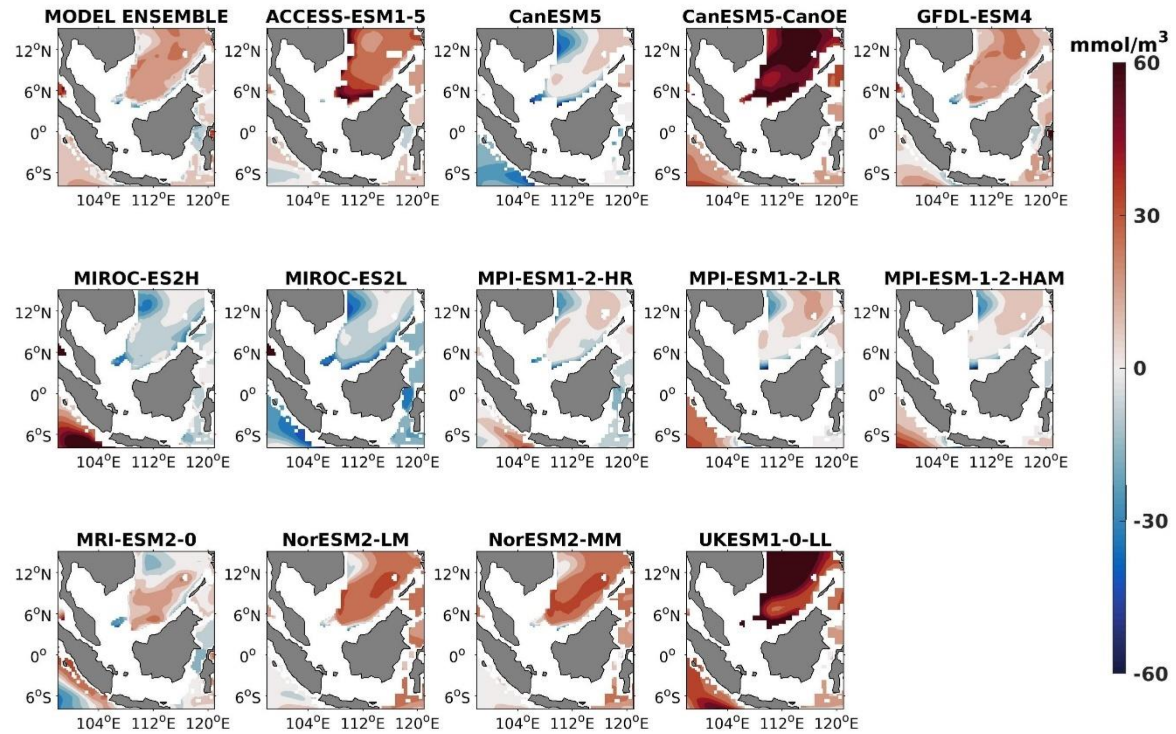

Figure 20. Same as Fig. 20 but for JJA.

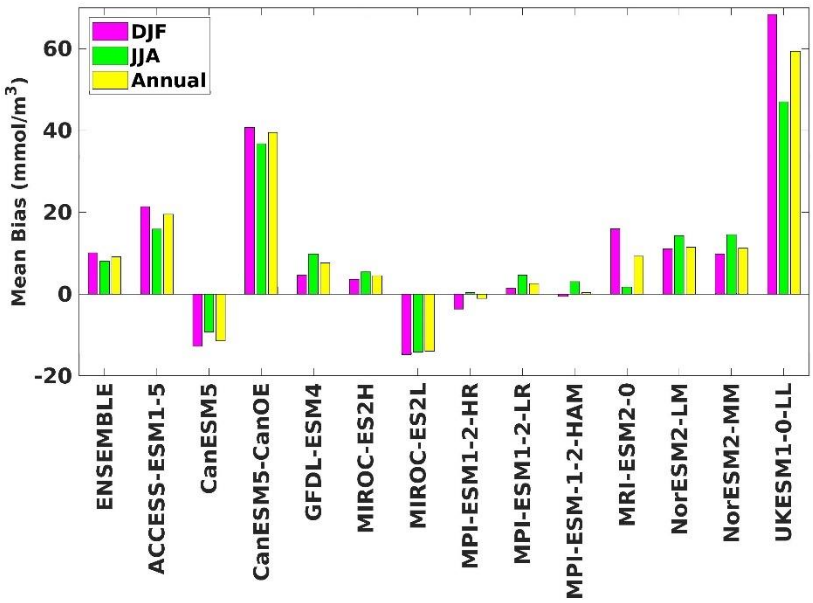

Figure 21. The mean bias of oxygen at 70-meter depth for both seasons (DJF, JJA) and annual.

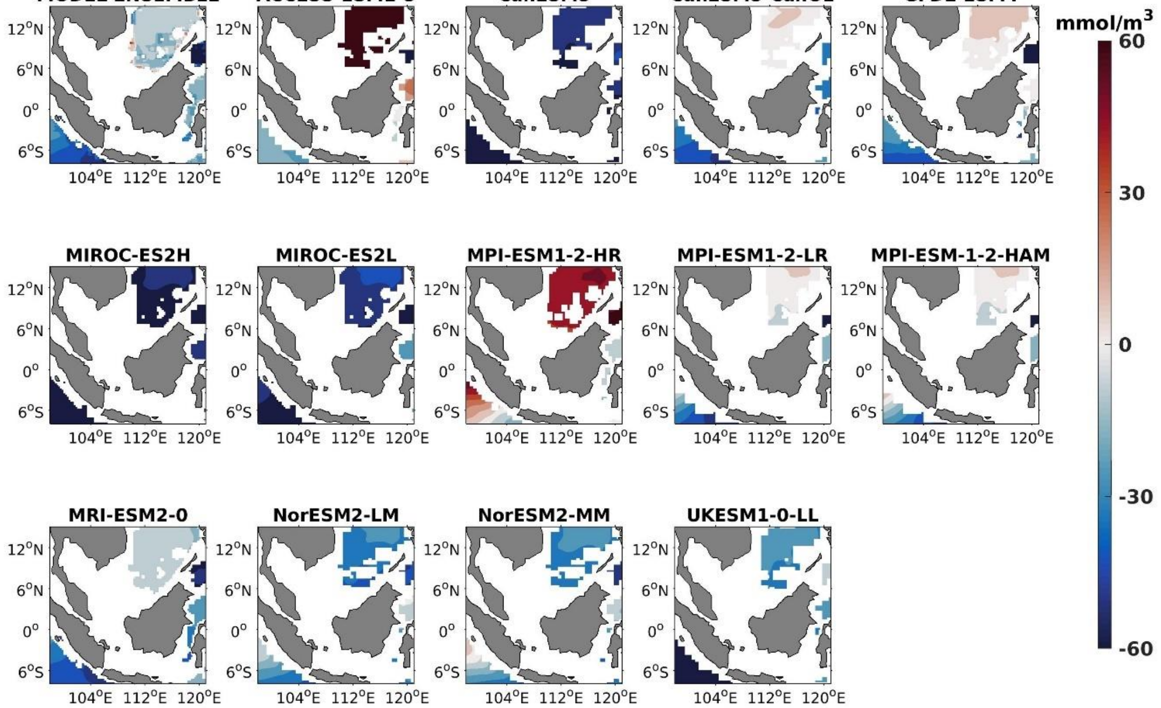

Figure 22. DJF spatial biases of oxygen at 1000-meter depth for 13 individual models and model ensemble relative to reference.

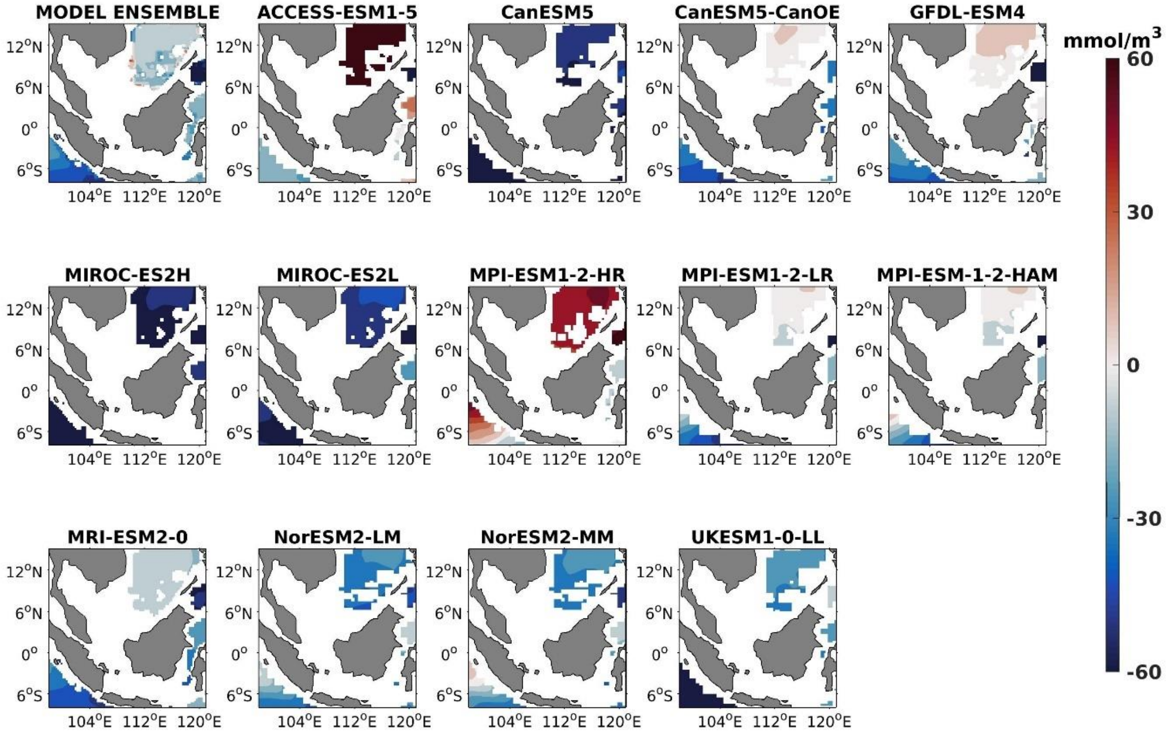

Figure 23. Same as Fig. 23 but for JJA.

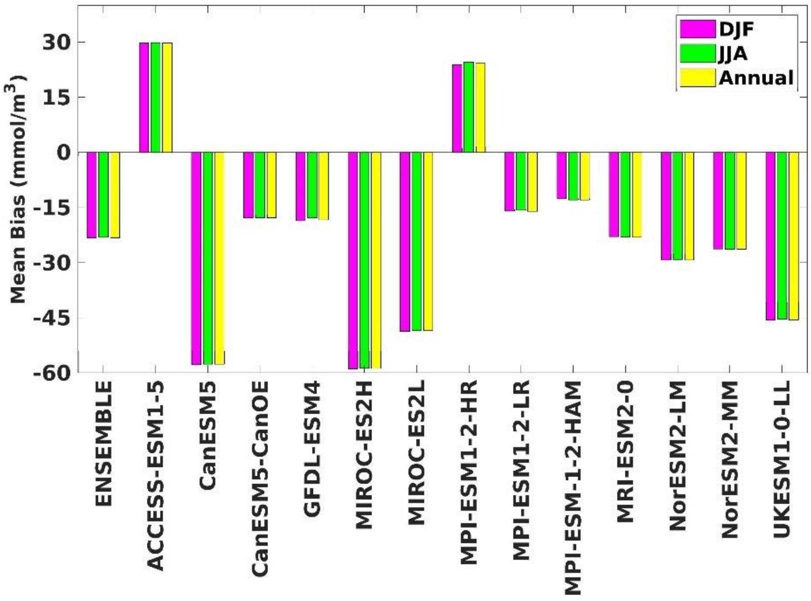

Figure 24. The mean bias of oxygen at 1000-meter depth for both seasons (DJF, JJA) and annual.

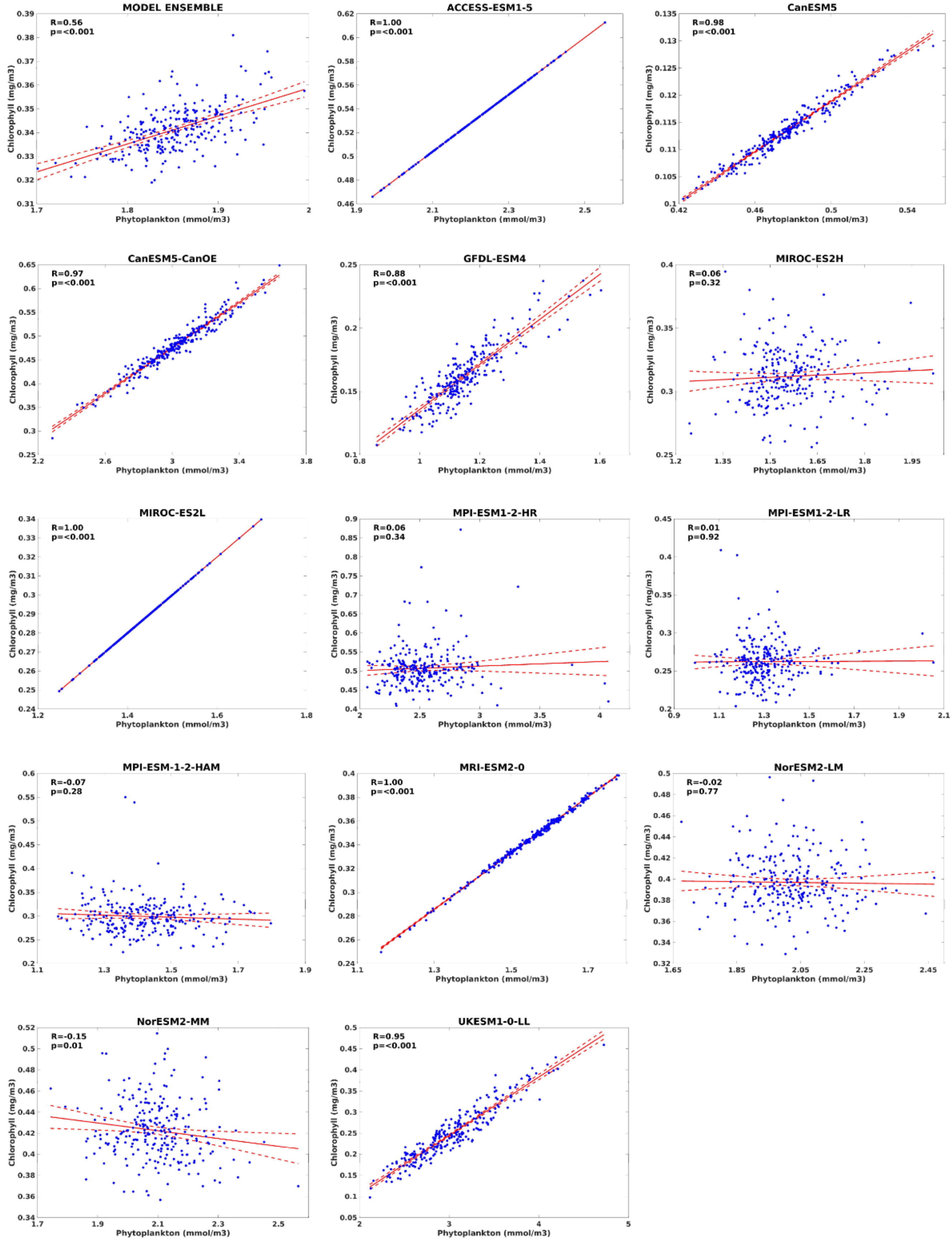

Figure 25. Relationships between chlorophyll and phytoplankton for model ensemble and 13 individual models in southern South China Sea during the study period (1993 – 2014). Dashed red lines represent 95% confidence interval with 0.05 as the level of significance (α).

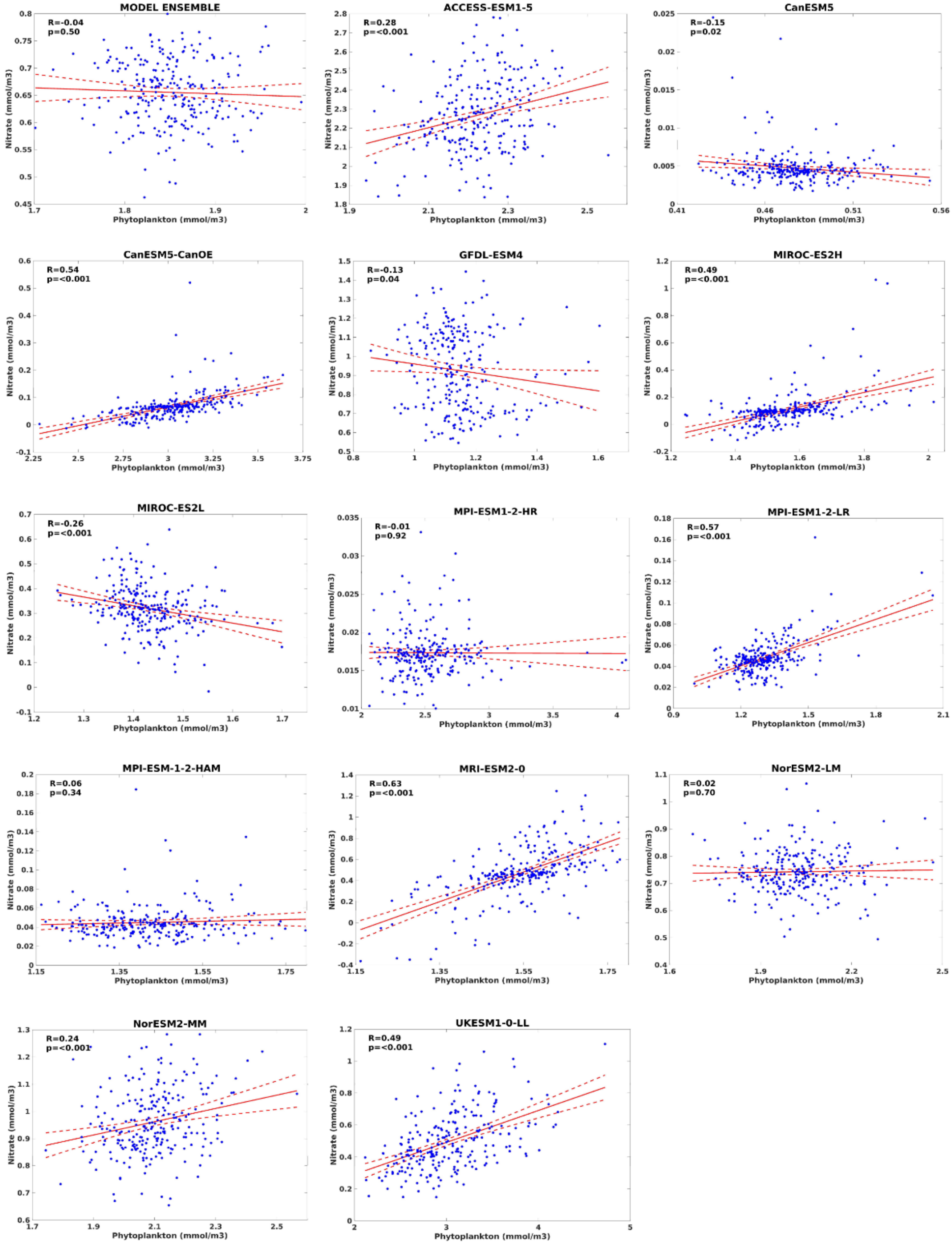

Figure 26. Relationships between nitrate and phytoplankton for model ensemble and 13 individual models in southern South China Sea during the study period (1993 – 2014). Dashed red lines represent 95% confidence interval with 0.05 as the level of significance (α).

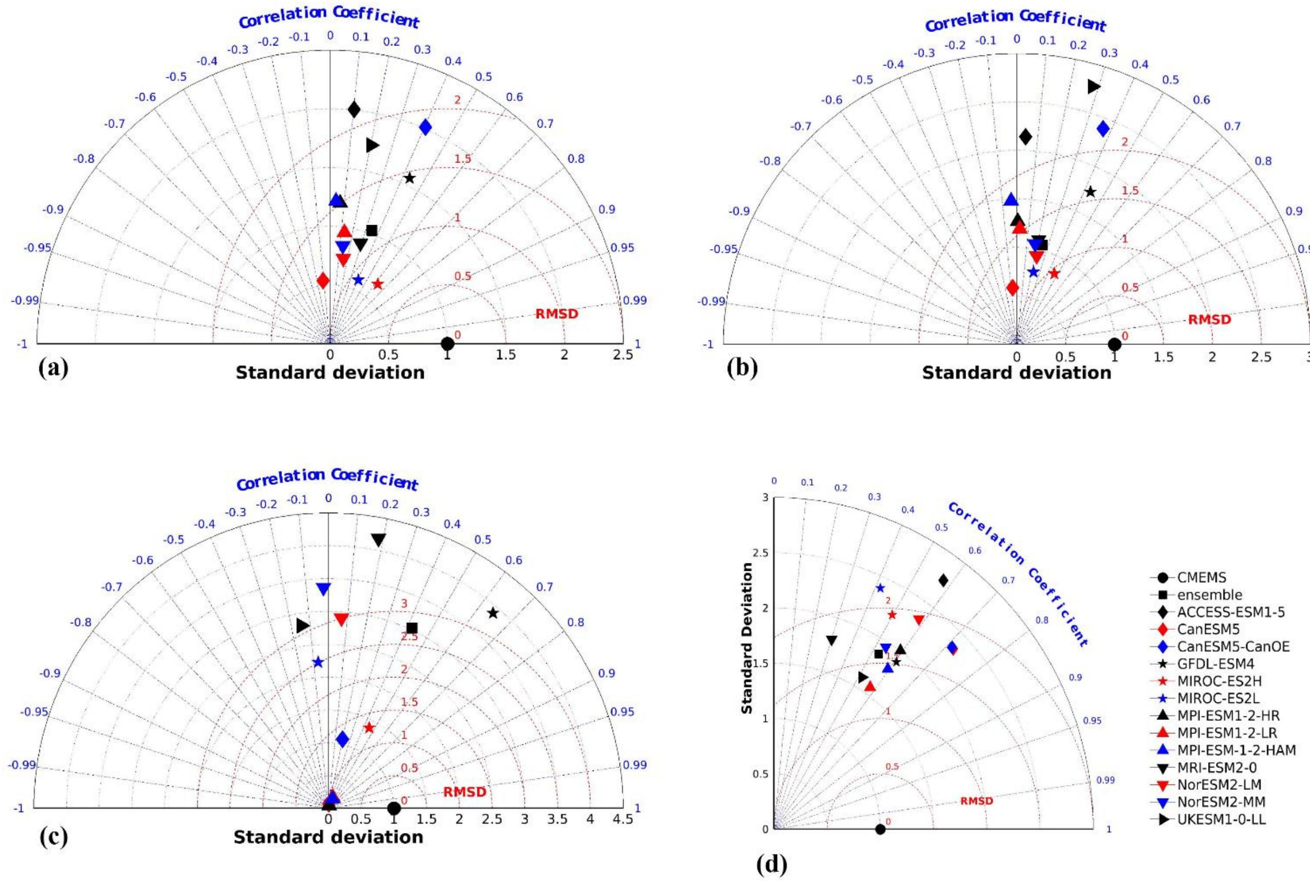

Figure 27. Annual Taylor Diagram for (a) chlorophyll, (b) phytoplankton, (c) nitrate and (d) oxygen.

Reference:

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., et al. (2006). Climate-driven trends in contemporary ocean productivity. *Nature*, *444*(7120), 752–755. https://doi.org/10.1038/nature05317

Chelton, D. B., Deszoeke, R. A., Schlax, M. G., El Naggar, K., & Siwertz, N. (1998). Geographical variability of the first baroclinic Rossby radius of deformation. *Journal of Physical Oceanography*, *28*(3). https://doi.org/10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2

Friedrichs, M. A. M., Hood, R. R., & Wiggert, J. D. (2006). Ecosystem model complexity versus physical forcing: Quantification of their relative impact with assimilated Arabian Sea data. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *53*(5–7). https://doi.org/10.1016/j.dsr2.2006.01.026

Glessmer, M. S., Oschlies, A., & Yool, A. (2008). Simulated impact of double-diffusive mixing on physical and biogeochemical upper ocean properties. *Journal of Geophysical Research: Oceans*, *113*(8). https://doi.org/10.1029/2007JC004455

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020a). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5), 2197–2244. https://doi.org/10.5194/gmd-13-2197-2020

Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020b). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, *13*(5). https://doi.org/10.5194/gmd-13-2197-2020

Jin, S., Wei, Z., Wang, D., & Xu, T. (2023). Simulated and projected SST of Asian marginal seas based on CMIP6 models. *Frontiers in Marine Science*, *10*. https://doi.org/10.3389/fmars.2023.1178974

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., et al. (2020). Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections. *Biogeosciences*, *17*(13), 3439–3470. https://doi.org/10.5194/bg-17-3439-2020

Laufkötter, C., John, J. G., Stock, C. A., & Dunne, J. P. (2017). Temperature and oxygen dependence of the remineralization of organic matter. *Global Biogeochemical Cycles*, *31*(7). https://doi.org/10.1002/2017GB005643

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019a). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4). https://doi.org/10.1029/2018MS001400

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019b). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018MS001400

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, *9*(1). https://doi.org/10.1002/2016MS000737

Petrik, C. M., Luo, J. Y., Heneghan, R. F., Everett, J. D., Harrison, C. S., & Richardson, A. J. (2022). Assessment and Constraint of Mesozooplankton in CMIP6 Earth System Models. *Global Biogeochemical Cycles*, *36*(11). https://doi.org/10.1029/2022GB007367

Séférian, R., Berthet, S., Yool, A., Palmiéri, J., Bopp, L., Tagliabue, A., et al. (2020). Tracking Improvement in Simulated Marine Biogeochemistry Between CMIP5 and CMIP6. *Current Climate Change Reports*. https://doi.org/10.1007/s40641-020-00160-0

Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and Evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4513–4558. https://doi.org/10.1029/2019MS001739

Sinha, B., Buitenhuis, E. T., Quéré, C. Le, & Anderson, T. R. (2010). Comparison of the emergent behavior of a complex ecosystem model in two ocean general circulation models. *Progress in Oceanography*, *84*(3–4). https://doi.org/10.1016/j.pocean.2009.10.003

Six, K. D., & Maier-Reimer, E. (1996). Effects of plankton dynamics on seasonal carbon fluxes in an ocean general circulation model. *Global Biogeochemical Cycles*, *10*(4). https://doi.org/10.1029/96GB02561

Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., et al. (2020). Ocean Biogeochemistry in GFDL's Earth System Model 4.1 and Its Response to Increasing Atmospheric CO2. *Journal of Advances in Modeling Earth Systems*, *12*(10). https://doi.org/10.1029/2019MS002043

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Tagliabue, A., Kwiatkowski, L., Bopp, L., Butenschön, M., Cheung, W., Lengaigne, M., & Vialard, J. (2021). Persistent Uncertainties in Ocean Net Primary Production Climate Change Projections at Regional Scales Raise Challenges for Assessing Impacts on Ecosystem Services. *Frontiers in Climate*, *3*. https://doi.org/10.3389/fclim.2021.738224

Thompson, B., Tkalich, P., Malanotte-Rizzoli, P., Fricot, B., & Mas, J. (2016). Dynamical and thermodynamical analysis of the South China Sea winter cold tongue. *Climate Dynamics*, *47*(5–6). https://doi.org/10.1007/s00382-015-2924-3

Tjiputra, J. F., Schwinger, J., Bentsen, M., Morée, A. L., Gao, S., Bethke, I., et al. (2020). Ocean biogeochemistry in the Norwegian Earth System Model version 2 (NorESM2). *Geoscientific Model Development*, *13*(5), 2393–2431. https://doi.org/10.5194/gmd-13-2393-2020

Wang, Z., Brickman, D., Greenan, B., Christian, J., DeTracey, B., & Gilbert, D. (2024). Assessment of Ocean Temperature Trends for the Scotian Shelf and Gulf of Maine Using 22 CMIP6 Earth System Models. *Atmosphere - Ocean*, *62*(1). https://doi.org/10.1080/07055900.2023.2264832